prediction model included age, gender, race (African-American non-Hispanic or otherwise), self-reported diabetes status, and self-reported hypertension status.

Multiple models were considered, including one logistic model predicting CKD Stage 1-4, another predicting CKD stage 3-4, and an ordered probit predicting stage of CKD. All models used the NHANES survey weights.

After model estimation, these parameter estimates were set aside.

Step 3: Assign CKD probabilities to BRFSS respondents

The Centers for Disease Control and Prevention annually performs a survey of adults that includes questions on factors, such as high blood pressure and diabetes, known to be associated with CKD. This survey, the Behavioral Risk Factor Surveillance System (BRFSS) is designed to be representative at the state level; North Carolina has roughly 17,000 respondents to the survey in any given year. Thus, we can apply the national prevalence estimates for CKD from NHANES to the NC population based on state specific health characteristics. For the final model used here, age, gender, history of hypertension, history of diabetes, and race of African-American were used as predictors.

The data from the 2005 BRFSS were formatted to correspond with the NHANES data. (High blood pressure awareness was not asked in the 2006 data, so 2005 was the most recent data available). With the data formatted in the same manner, predictions can be generated using the average associations between observed factors and CKD stage at a national level. A simplified example may be helpful in illustrating this approach. For purposes of the example, assume that a survey reveals that 20 percent of males and 60 percent of females have a certain characteristic. If a group of similar people is 50 percent male and 50 percent female, one estimate is that .5 x 20% + .5 x 60% = 40% of the group has the characteristic. If a second group is 80 percent male and 20 percent female, an estimate would be .8 x 20% + .2 x 60% = 28%. Assuming that the relationships in the development (here, NHANES) dataset is similar to the relationships in the *estimation* (here, NC using BRFSS) population, then we are able to estimate a valid prevalence rate of CKD in NC using this approach.[3]

Table A-1 presents 5 different sets of estimates. Column A is the prevalence as presented in Coresh using 1999-2004 data. Column B is the NC IOM's analysis based on 1999-2004. Both these are standardized to the 2000 standard population. Overall, the predictions are quite similar between the Coresh et al model analysis and the replication by the Task Force. Column C uses the weights in the NHANES data, meaning the predictions apply to a 1999-2004 population. This slightly lowers the estimated prevalence of CKD. Column D applies the estimated relationship in NHANES and looks at the in-sample predictive power. The in-sample predictive power, as expected, is overall quite good – the predicted prevalence is very similar to the estimated